

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-259175

(43)Date of publication of application : 22.09.2000

(51)Int.Cl.

G10L 15/18
G10L 15/14

(21)Application number : 11-060418

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 08.03.1999

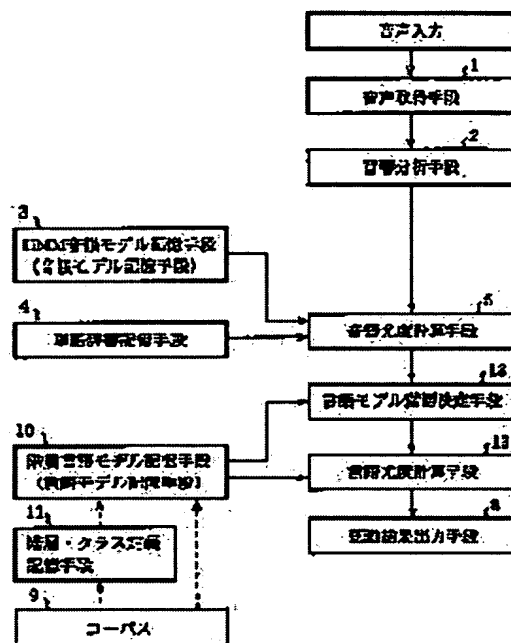
(72)Inventor : MARUTA YUZO

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain a voice recognition device having a language model which gives zero or a very small value as a bigram probability for an abnormal word chain even though the amount of corpus is not sufficient.

SOLUTION: The device is provided with a language model hierarchy determining means 12 which has a statistical language model having a hierarchical structure and determines an optimum hierarchy to compute a language bigram by a preceding word, and a language likelihood computing means 13 which gives a very small value to the language likelihood of a strange word chain while considering the frequency of the appearance of an optimum hierarchy class for the bigram probability of the chain word. Then, zero or a small value close to zero is given as the probability for an abnormal word chain to improve the recognition performance/processing speed.



LEGAL STATUS

[Date of request for examination]

21.09.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

This Page Blank (uspto)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-259175

(P2000-259175A)

(43) 公開日 平成12年9月22日 (2000.9.22)

(51) Int.Cl.⁷

識別記号

F I

テームコード* (参考)

G 1 0 L 15/18

G 1 0 L 3/00

5 3 7 D 5 D 0 1 5

15/14

5 3 5 Z

審査請求 未請求 請求項の数9 O L (全 13 頁)

(21) 出願番号 特願平11-60418

(22) 出願日 平成11年3月8日 (1999.3.8)

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72) 発明者 丸田 裕三

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(74) 代理人 100066474

弁理士 田澤 博昭 (外1名)

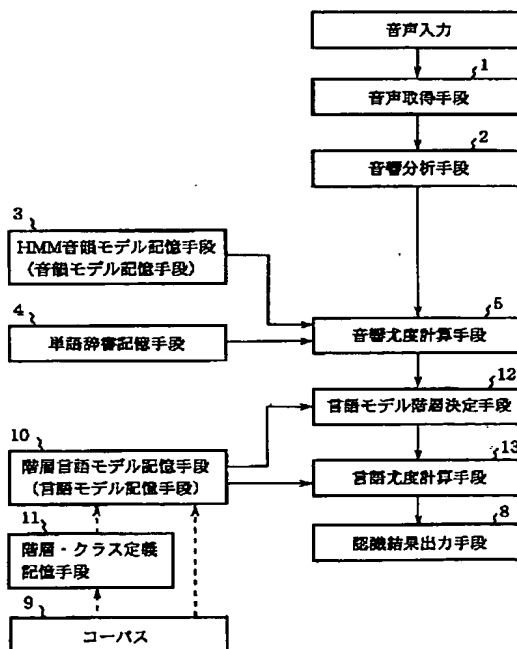
Fターム(参考) 5D015 HH23 JJ07

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 コーバスの量が十分でない場合でも、異常な単語連鎖についてはバイグラム確率として0ないしきわめて小さい値を与える言語モデルをもった音声認識装置を提供する。

【解決手段】 階層構造の統計的言語モデルをもち、先行単語によって、言語バイグラムを計算するための最適な階層を決定する言語モデル階層決定手段12、および連鎖単語のバイグラム確率を最適な階層のクラスの出現頻度を考慮することによって、奇異な単語連鎖の言語尤度についてきわめて小さい値を与える言語尤度計算手段13を備え、異常な単語連鎖についてはバイグラム確率として0ないし0に近い小さい値を与えることを可能として、認識性能・認識処理速度を向上させる。



【特許請求の範囲】

【請求項1】 コーバスから作成した統計的言語モデルを用いて入力音声の音声認識を行う音声認識装置において、

入力音声をデジタルデータ化し、それを音声データとして記憶する音声取得手段と、

前記音声データを所定時刻ごとに音響分析して音響特徴ベクトルを出力する音響分析手段と、

音韻モデルを記憶する音韻モデル記憶手段と、

単語辞書を記憶する単語辞書記憶手段と、

前記音響分析手段より出力された音響特徴ベクトル、前記音韻モデル記憶手段に記憶されている音韻モデル、および前記単語辞書に記憶されている単語辞書から、認識仮説の音響尤度を計算する音響尤度計算手段と、

前記統計的言語モデルを記憶する言語モデル記憶手段と、

単語遷移した前記認識仮説について、その単語連鎖の言語モデルを計算する際の言語モデルの階層を決定する言語モデル階層決定手段と、

前記言語モデル記憶手段に記憶されている統計的言語モデルを参照し、前記言語モデル階層決定手段の決定した言語モデルの階層から認識仮説の言語尤度を計算する言語尤度計算手段と、

前記音響尤度計算手段の計算した認識仮説の音響尤度と、前記言語尤度計算手段の計算した認識仮説の言語尤度に基づいた、最終的な認識候補を出力する認識結果出力手段とを備え、

前記統計的言語モデルは階層構造をもち、音声認識の処理状況に応じて適切な階層の言語モデルを使用することを特徴とする音声認識装置。

【請求項2】 統計的言語モデルは、各先行単語によって最適な階層の言語モデルを使用することを特徴とする請求項1記載の音声認識装置。

【請求項3】 統計的言語モデルは、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーバスの出現頻度が0の場合には、その単語バイグラムの確率値を0または0に近い小さい正值とすることを特徴とする請求項2記載の音声認識装置。

【請求項4】 統計的言語モデルは、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーバスの出現頻度が0より大きく、かつ、コーバス内の単語バイグラムの出現頻度が0である場合には、その単語バイグラムの確率値を0より大きい値に設定することを特徴とする請求項2記載の音声認識装置。

【請求項5】 統計的言語モデルにおける先行単語に使用する最適な階層として、所属するクラスの単語バイグラムの出現頻度が0でない後続単語の数の変化率が最大になる直前の階層を選択することを特徴とする請求項2記載の音声認識装置。

【請求項6】 統計的言語モデルは、階層構造をもつクラスを、コーバス以外の外部知識を用いて作成することを特徴とする請求項1記載の音声認識装置。

【請求項7】 統計的言語モデルは、階層構造をもつクラスを、外部知識を用いず、各階層においてそれぞれクラスターリングをすることによって作成することを特徴とする請求項1記載の音声認識装置。

【請求項8】 統計的言語モデルは、階層構造をもつクラスを、ある階層については、コーバス以外の外部知識を用いて作成し、別の階層については外部知識を用いずにクラスターリングをすることによって作成することを特徴とする請求項1記載の音声認識装置。

【請求項9】 統計的言語モデルは、クラスターリングをすべき単語を行列表現して、行と列の入れ替えによってコーバスの出現頻度が0でない部分を局在させ、その部分に外接するような部分行列をクラスとすることを特徴とする請求項1記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声認識装置に関し、特に、統計的言語モデルを用いた音声認識装置に関するものである。

【0002】

【従来の技術】図15は「新聞記事を用いた大語彙連続音声認識の検討」（電子情報通信学会技術研究報告、Vol. 95, No. 430, pp 63～pp 68, 1995年12月14日）に開示された、統計的言語モデルを用いた従来の音声認識装置の概略構成を示すブロック図である。図において、1は音声取得手段、2は音響分析手段、3はHMM（隠れマルコフモデル）音韻モデル記憶手段、4は単語辞書記憶手段、5は音響尤度計算手段、6は言語モデル記憶手段、7は言語尤度計算手段、8は認識結果出力手段、9はコーバスである。

【0003】次に動作について説明する。ここで、図12は単語辞書の具体例を示す説明図、図13は単語のHMMの構造の一例を示す説明図であり、図14は時刻が進むにつれて認識仮説が展開される状況を示す説明図である。また、図16は図15に示した従来の音声認識装置の動作の概略を示したフローチャートである。

【0004】図16のフローチャートに示した動作を開始するに先立って、まず、HMM音韻モデル記憶手段3が外部記憶装置（図示省略）からHMM音韻モデルのデータを読み込んで記憶する。単語辞書記憶手段4も同様に、上記外部記憶装置から単語辞書を読み込んで記憶する。この単語辞書は図12に示すように、各単語についてそれぞれ漢字表記、ひらがな表記、音素表記のデータをもっている。

【0005】さらに単語辞書記憶手段4は、各単語の音素表記からその単語のHMM音響モデルを生成する。図13はHMM音響モデルの構成例を示したものであり、

図示のように単語はそれぞれ各音素に分解され、各音素ごとに対応するHMM音韻モデルを当てはめて、それらを連結することにより、各単語のHMM音響モデルを構成する。

【0006】次に音声取得手段1は、図16のステップST1において、入力された音声をA/D変換し、デジタルデータ化された音声データを記憶する。次に制御手段(図示省略)は、ステップST2において時刻 t を初期値として $t=0$ に設定する。次に音声取得手段1は、ステップST3においてその時刻 t の音声データを取り込み、音響分析手段2はステップST4において、取り込まれた音声データを音響分析して音響特徴ベクトルを計算する。次に、音響尤度計算手段5はステップST5において、音響特徴ベクトル、単語辞書、およびHMM音韻モデルデータから、各認識仮説の対数音響尤度を計算して認識仮説を展開する。

【0007】図14は各時刻における認識仮説の展開の様子を示しており、ここでは、簡略化のために各音素のHMM状態数は1としている。図において、それぞれの四角が認識仮説であり、各認識仮説は認識している単語と現在の音素、対数音響尤度、対数言語尤度、対数総合尤度を情報として持っている。例えば、認識仮説の認識単語を「記事」とすると、フレーム(時刻)が進行するにつれて、HMM音響モデルが自己ループして音素が進行しない認識仮説と、HMM音響モデルが進行して音素が進行する認識仮説に展開され、認識仮説の数が増えてゆくことになる。特に「記事」の終端の音素/i/が終了すると(図14の太枠の認識仮説)次の認識単語に移すが、「記事」の次には「銀」、「議員」、「議 *

*会」、…とさまざまな単語が遷移し得るため、そのおのに対して独立に認識仮説を割り当てる。

【0008】次に、言語尤度計算手段7はステップST6において、単語遷移を行った直後の認識仮説について、言語モデル記憶手段6に記憶されている統計的言語モデルをもとに言語尤度を計算し、その対数値である対数言語尤度に、重みを付けを行った対数音響尤度を加算することによって対数総合尤度を計算する。なお、この言語尤度の詳細については後に説明する。

【0009】次に、図示を省略した制御手段はステップST7において、音声入力が終了したか否かを検出し、終了していない場合にはステップST8において、時刻 t を $t+1$ にインクリメントして処理をステップST3に戻し、上記ステップST3からステップST6までの処理を音声入力が終了するまで繰り返す。ステップST7にて音声入力の終了が検出されると、認識結果出力手段8はステップST9において、発声的全区間において計算が終わった認識仮説について、その対数総合尤度の大きい順に、認識結果として認識候補(単語列)を出力する。

【0010】ここで、その言語尤度の計算の詳細について説明する。言語尤度は統計的言語モデルが用いられることが多く、特にNグラムモデルを用いる場合が多い。Nグラムモデルでは、単語列 $W=w_1 w_2 \dots w_n$ の言語尤度 $P(w_1 w_2 \dots w_n)$ として、条件付き確率を用いて、以下に示した式(1)を与える。

【0011】

【数1】

$$P(w_1 w_2 \dots w_n) = \prod_{i=1}^n P(w_i | w_{i-1} w_{i-2} \dots w_{i-N+1}) \quad \dots (1)$$

【0012】なお、上記式(1)で $N=2$ のときバイグラム、 $N=3$ のときトライグラムという。式(1)において、 w_i や w_{i-1} などには文頭・文末を示す特別な記号を割り当てることが多い。以下では簡略化のため、 $N=2$ のバイグラムモデルを説明のために用いるが、 $N>2$

※2の場合でも同様である。なお、バイグラムモデルの場合、言語尤度は次の式(2)のようになる。

【0013】

【数2】

$$\begin{aligned} P(w_1 w_2 \dots w_n) &= \prod_{i=1}^n P(w_i | w_{i-1}) \\ &= P(w_1 | w_0) P(w_2 | w_1) \dots \quad \dots (2) \end{aligned}$$

【0014】ところで、これらの条件付き確率は、一般にコーパスと呼ばれる大量の文書から計算される。しかしながら、文字の組み合わせは一般に膨大になるため、コーパスで統計上十分である場合は少ない。通常、 α ★

★のことをコーパス9のスパース性という。このような場合、以下に示す式(3)のようにスムージングをする場合が多い。

【0015】

$$\begin{aligned} P(w | x) &= (1 - \alpha) \cdot N(x, w) / N(x) \\ &\quad (N(x, w) > 0 \text{ の場合}) \\ P(w | x) &= \alpha \cdot P(w) \quad (N(x, w) = 0 \text{ の場合}) \\ &\quad \dots (3) \end{aligned}$$

【0016】ここで、上記式(3)において、 $N(x, w)$ はコーパス9内に存在する単語 $x-w$ の連鎖の数、 $N(x)$ はコーパス9内に存在する単語 x の数、 $P(w)$ は単語 w のユニグラム確率、 α はスムージングのための係数である。なお、 α はコーパス9が統計上十分になるにしたがって $\alpha \rightarrow 0$ に収束する。

【0017】また、コーパス9のスパース性を回避する*

$$\begin{aligned} P(w|x) &= P(C|x) \cdot P(D|C) \cdot P(w|D) \\ P(D|C) &= N(d, c) / N(c) \\ P(w|D) &= N(w) / N(d) \\ P(C|x) &= 1 \end{aligned}$$

【0019】なお、このような従来の音声認識装置に関連のある記載がなされている文献としては、この他にも、例えば、特開平8-254990号公報、特開平97535号公報、特開平8-329080号公報などがある。

【0020】

【発明が解決しようとする課題】「電子情報通信学会技術研究報告」に開示された、統計的言語モデルを用いた従来の音声認識装置は以上のように構成されているので、特にコーパス9の量が十分でない場合には α が無視できなくなり、結果として、人間が奇異に感じる単語連※

$$P(\text{“銀”} | \text{“国会”}) = \alpha \cdot P(\text{“銀”}) > 0 \quad \dots (5)$$

【0022】一方、品詞バイグラムを用いた場合でも、「銀」、「国会」はともに名詞なので、 $P(\text{“銀”} | \text{“国会”})$ は次の式(6)に示すようになり、同様に言★

$$P(\text{“銀”} | \text{“国会”}) = P(\text{“名詞”} | \text{“国会”}) \cdot P(\text{“名詞”} | \text{“名詞”}) \cdot P(\text{“銀”} | \text{“名詞”}) > 0$$

【0024】この発明は上記のような課題を解決するためになされたもので、コーパスの量が十分でない場合でも、異常な単語連鎖については、バイグラムの確率値として0ないしは0に近い小さな値を与える言語モデルを用いた音声認識装置を得ることを目的とする。

【0025】

【課題を解決するための手段】この発明に係る音声認識装置は、入力音声デジタルデータ化した音声データを所定時刻ごとに音響分析して音響特徴ベクトルを生成し、この音響特徴ベクトルと、音韻モデル記憶手段の記憶する音韻モデルおよび単語辞書の記憶する単語辞書とから認識仮説の音響尤度を計算し、また、言語モデル記憶手段の記憶する統計的言語モデルを参照して、言語モデル階層決定手段の決定した言語モデルの階層から認識仮説の言語尤度を計算し、それら認識仮説の音響尤度と言語尤度に基づいて最終的な認識候補を出力するとともに、その統計的言語モデルに階層構造をもたせ、音声認識の処理状況に応じて適切な階層の言語モデルを使用するようにしたものである。

【0026】この発明に係る音声認識装置は、各先行単

*ために、品詞バイグラムなどの言語モデルを用いる場合がある。この場合、単語 x, w はそれぞれの品詞 C, D に属するものとする、次の式(4)が与えられる。ただし、この式(4)中、単語 c は品詞 C に属する単語について、単語 d は品詞 D に属する単語について、そのすべてをとる。

【0018】

..... (4)

※鎖でも、バイグラムの確率値は0でない値をとってしまうことがある。例えば、国会-銀という単語対は、通常のコーパス9には存在せず、かつ一般には考えにくいので、 $P(\text{“銀”} | \text{“国会”}) = 0$ であるのが妥当であるが、式(3)に従うと、次の式(5)に示すものとなって、「国会銀」という認識仮説にも言語尤度としてある値が与えられてしまうため、「国会議員」と発声した場合でも、発声の曖昧さによっては「国会銀」と誤認識してしまう場合があるという課題があった。

【0021】

★語モデルで小さい言語尤度を与えるのは困難になるという課題があった。

【0023】

..... (6)

語によって最適な階層の言語モデルを使用するようにしたものである。

【0027】この発明に係る音声認識装置は、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーパスの出現頻度が0である場合には、その単語バイグラムの確率値を0またはきわめて小さい正值とするようにしたものである。

【0028】この発明に係る音声認識装置は、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーパスの出現頻度が0より大きく、かつ、コーパス内の単語バイグラムの出現頻度が0の場合には、その単語バイグラムの確率値を0より大きい値に設定するようにしたものである。

【0029】この発明に係る音声認識装置は、統計的言語モデルにおける先行単語に使用する最適な階層として、所属するクラスの単語バイグラムの出現頻度が0でない後続単語の数の変化率が最大になる直前の階層を選択するようにしたものである。

【0030】この発明に係る音声認識装置は、コーパス以外の外部知識を用いて階層構造をもつクラスの作成を

行うようにしたものである。

【0031】この発明に係る音声認識装置は、外部知識を用いず、各階層においてそれぞれクラスタリングをすることによって、階層構造をもつクラスの作成を行うようにしたものである

【0032】この発明に係る音声認識装置は、ある階層についてはコーパス以外の外部知識を用いて、別の階層については外部知識を用いずクラスタリングをすることによって、階層構造をもつクラスの作成を行うようにしたものである。

【0033】この発明に係る音声認識装置は、クラスタリングすべき単語を行列表現してその行と列とを入れ替えることによって、コーパスの出現頻度が0でない部分を局在させ、その部分に外接するような部分行列をクラスとするようにしたものである。

【0034】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態1. 図1はこの発明の実施の形態1による音声認識装置の概略構成を示すブロック図である。図において、1は入力音声デジタルデータ化し、それを音声データとして記憶する音声取得手段であり、2はこの音声取得手段1において取得された音声データを所定時刻ごとに音響分析し、音響特徴ベクトルを出力する音響分析手段である。3は図示を省略した外部記憶装置からHMM音韻モデルを読み込み、そのHMM音韻モデルを記憶する音韻モデル記憶手段としてのHMM音韻モデル記憶手段であり、4は同じく図示を省略した外部記憶装置から単語辞書を読み込み、その単語辞書を記憶する単語辞書記憶手段である。5は音響分析手段2より出力される音響特徴ベクトルと、HMM音韻モデル記憶手段3に記憶されているHMM音韻モデル、および単語辞書記憶手段4に記憶されている単語辞書から、各認識仮説についての音響尤度を計算する音響尤度計算手段である。8は各認識仮説の音響尤度と言語尤度とに基づいた、最終的な認識候補を出力する認識結果出力手段である。9はその言語尤度を計算するための条件付き確率を求める際に用いられる、大量の文書によるコーパスである。

【0035】なお、これら音声取得手段1、音響分析手段2、HMM音韻モデル記憶手段3、単語辞書記憶手段4、音響尤度計算手段5、認識結果出力手段8、コーパス9は、図15に同一符号を付して示した従来の音声認識装置におけるそれらに相当する部分である。

【0036】10は統計的言語モデルを記憶する言語モデル記憶手段としての階層言語モデル記憶手段であり、コーパス9と後述する階層・クラス定義記憶手段に記憶されている階層・クラス定義を用いて作成した、階層構造をもつ統計的言語モデルを記憶している点で、図15に符号6を付して示した言語モデル記憶手段とは異なっている。なお、この階層言語モデル記憶手段10に記憶

されている統計的言語モデルは、認識処理の状況に応じて適切な階層の言語モデルが使用される。11は階層言語モデル記憶手段10に記憶されている階層構造をもった統計的言語モデルについて、その階層・クラスの定義を記憶する階層・クラス定義記憶手段である。12は単語遷移した認識仮説について、その単語連鎖の言語モデルを計算する際の言語モデルの階層を決定する言語モデル階層決定手段である。13は階層言語モデル記憶手段10に記憶されている統計的言語モデルを参照して、認識仮説の言語尤度を計算して、その対数値である対数言語尤度と、音響尤度計算手段5によって計算された音響尤度の対数値である対数音響尤度より、対数総合尤度を計算する言語尤度計算手段であり、言語モデル階層決定手段12によって決定された階層の言語モデルから言語尤度の計算を行い、対数言語尤度および対数総合尤度を計算している点で、図15に符号7を付して示した従来の言語尤度計算手段とは異なっている。

【0037】次に動作について説明する。ここで、図2はこの実施の形態1における認識処理の概略を示すフローチャート、図3は階層・クラス定義の一例を示す説明図、図4は階層構造を有する言語モデルの、先行単語xと後続単語wの各階層のコーパス9内での出現頻度を示す説明図であり、図5は言語モデル階層決定手段12の概略動作を示すフローチャート、図6はそれによって計算された配列をグラフで示した説明図である。また、図12は単語辞書の具体例を示した説明図、図13は単語のHMMの構造の一例を示す説明図であり、図14は時刻が進むにつれて認識仮説が展開される状況を示す説明図である。

【0038】まず最初に、図3および図4を参照しながら、この実施の形態1で用いられている階層的な言語モデルについて説明する。図3は先験的な知識を用いて単語を階層化した階層・クラス定義の一例である。この場合には、単語を5つの階層に分け、最上層は品詞、最下層は単語として、中間層は「医療-薬品-塗布薬」などの先験的な知識を用いてそれぞれの上位の階層を分類している。なお、これら各階層には、最上層から最下層に対して、C4、C3、…、C0のクラスレベルが付与されている。階層・クラス定義記憶手段11はこのような階層・クラス定義を記憶している。

【0039】次に、この階層・クラス定義記憶手段11に記憶されている上記階層・クラス定義と、コーパス9とを参照して、各階層についてのコーパス9内の単語連鎖の頻度を数値化する。図4ではx-wの単語連鎖がコーパス9内で5回計数された($N(x, w) = 5$)ことを示している。階層言語モデル記憶手段10はこれら各階層の単語連鎖の頻度を先行単語xごとに計数して記憶しておく。

【0040】次いで、この実施の形態1による音声認識装置の認識処理動作を図2のフローチャートにしたがっ

て説明する。この図2のフローチャートに示した処理動作を開始するに先立って、まず、HMM音韻モデル記憶手段3が、図示を省略した外部記憶装置からHMM音韻モデルのデータを読み込み、そのHMM音韻モデルを記憶する。同様に、単語辞書記憶手段4は図示を省略した上記外部記憶装置から単語辞書を読み込み、その単語辞書を記憶する。ここで、この単語辞書は、例えば、図12に示すような構成になっており、各単語についてそれぞれ漢字表記、ひらがな表記、音素表記のデータをもっている。さらに、この単語辞書記憶手段4は各単語の音素表記からその単語のHMM音響モデルを生成する。

【0041】図13はこのHMM音響モデルの構成例を単語「議会」について示したものである。図示のように単語「議会」はそれぞれ、/g/, /i/, /k/, /a/, /i/の5つの音素に分解され、各音素ごとに対応するHMM音韻モデルを当てはめてそれらを連結することにより、当該単語「議会」のHMMの音響モデルを構成する。なお、他の単語についてもこれと同様にしてHMM音響モデルが構成される。

【0042】その後、音声取得手段1はまず、ステップST10において、入力された音声をA/D変換してデジタルデータ化し、それを音声データとして記憶する。次に、図示を省略した制御手段が、ステップST11において時刻tを初期値としてt=0に設定する。次に、音響分析手段2はステップST12において、音声取得手段1が取得した時刻t（この場合t=0）における音声データを取り込み、ステップST13において、その取り込んだ時刻tの音声データを音響分析して、音響特徴ベクトルを計算する。

【0043】次に、音響尤度計算手段5はステップST14において、単語辞書記憶手段4に記憶されている単語辞書から認識仮説を作成し、音響分析手段2が計算した音響特徴ベクトルとHMM音韻モデル記憶手段3に記憶されているHMM音韻モデルデータから、各認識仮説の対数音響尤度を計算して認識仮説を展開する。なお、この対数音響尤度の計算方法は、例えば特公平4-22276号公報などにも開示されている周知のものであるため、ここではその説明は割愛する。 *

$$N(x, C) \equiv \sum_{w \in C} N(x, w)$$

【0049】判定の結果、 $N(x, C) \neq 0$ であった場合はステップST25において、配列G(i)、すなわち、所属するクラス頻度が0でない単語の数に1を加える。次にステップST26において、各階層iにおける全単語の終了を検出するまで、ステップST27でjをインクリメントしながら処理をステップST23に戻し、さらにステップST28において全階層の終了を検出するまで、ステップST29でiをインクリメントし

*【0044】図14には各時刻（フレーム）における認識仮説の展開の様子が示されている。なお、簡略化のために、ここでは各音素のHMM状態数は1としている。図において、それぞれの四角が認識仮説を表しており、各認識仮説は認識している単語と現在の音素、対数音響尤度、対数言語尤度、対数総合尤度を情報としてもっている。例えば、認識仮説の認識単語を「記事」とすると、フレームがt=0から1, 2, ...と進行するにつれて、HMM音響モデルが自己ループして音素が進行しない認識仮説と、HMM音響モデルが進行して音素が進行する認識仮説に展開され、認識仮説の数が増えてゆくことになる。特に「記事」の終端の音素/i/が終了すると（図14の太枠の認識仮説）次の認識単語に遷移するが、「記事」の次には「銀」、「議員」、「議会」、...とさまざまな単語が遷移し得るため、そのおのおのに対して独立に認識仮説を割り当てる。なお、ここまでの動作は従来の音声認識装置の場合と同様である。

【0045】次に、言語モデル階層決定手段12はステップST15において、各認識候補にて単語が遷移する際に、その単語についての最適な階層を決定する。すなわち、認識仮説が単語遷移をして、認識候補における先行単語がxの場合、どの階層の言語モデルを使用するかを決定する。

【0046】図5のフローチャートにその言語モデル階層決定手段12による決定手続きの動作を示す。まずステップST20において階層iをi=0に設定する。次にステップST21において配列G(i)をG(i)=0に初期化する。次にステップST22において単語番号jをj=0に設定する。次にステップST23において単語w_iが属する階層iのクラスCを求める。

【0047】次に、得られた階層iのクラスCについて、先行単語xとクラスCに属する単語wの単語連鎖のコーパス9内の出現頻度N(x, C)を、階層言語モデル記憶手段10に記憶されている統計的言語モデルから求め、ステップST24においてそれが0であるか否かの判定を行う。なお、このN(x, C)は次に示した式(7)によって定義される。

【0048】

【数3】

$$\dots (7)$$

ながら処理をステップST21に戻すことによって、上記処理を各単語w_iおよび各階層iについて行う。

【0050】以上の処理が終わった段階で配列G(i)をグラフにしてみると、階層iが上がるにつれてG(i)は増加するので、例えば図6に示すような右肩上がりのカーブになる。この例では、階層2から階層3に移る際に配列G(i)が急増している。これは、階層2では単語が属しているクラスの頻度が0である場合が多

11

かったのに対して、階層3では単語が属しているクラスの頻度が0の場合は減少している（この場合は0になっている）ということであり、したがって階層3ではクラスの分割が粗すぎることを示している。そこでステップST30において、 $G(i+1) - G(i)$ が最大となるような階層（この場合には階層2）を最適な階層と決定する。

【0051】この場合、階層2内のクラスで、 $N(x, C) = 0$ であるクラスCは、コーパス9のスパース性によってたまたま連鎖が出現していないのではなく、言語的に出現していないと考えることができる。したがって、先行単語xとこのクラスCに属する単語の連鎖はないと考えてよい。

$$P(w|x) = (1 - \alpha) \cdot N(x, w) / N(x) \\ N(x, C) > 0 \text{ かつ } N(x, w) > 0 \text{ の場合} \\ P(w|x) = \alpha \cdot P(w) \\ N(x, C) > 0 \text{ かつ } N(x, w) = 0 \text{ の場合} \\ P(w|x) = 0 \quad N(x, C) = 0 \text{ の場合}$$

【0054】次に従来の音声認識装置と同様に、図示を省略した制御手段はステップST17において、音声入力力が終了したか否かを検出する。その結果、音声入力力が終了していない場合には、ステップST18において、時刻tをt+1にインクリメントして処理をステップST12に戻し、上記ステップST12からステップST16までの処理を、ステップST17で音声入力の終了が検出されるまで繰り返す。ステップST17にて音声入力の終了が検出されると、結果出力手段8はステップST19において、発声の全区間において計算が終わった認識仮説について、その対数総合尤度の大きい順に、認識結果として、認識候補（単語列）を出力する。

【0055】以上のように、この実施の形態1によれば、言語モデルを階層構造をもつ統計的言語モデルを用いて、先行単語により最適な階層を決定し、後続単語が属するクラスの、コーパス9内での計数が0のときにはバイグラム確率を0としているので、従来のバックオフ法による音声認識装置では実現できなかった、言語的につながり得ない奇異な単語連鎖に対して、その対数総合尤度をきわめて小さくすることができるようになるため、認識性能の向上をはかることができるという効果が得られる。

【0056】また、フレームの途中で不要な認識仮説を棄却する、いわゆる枝刈り処理を行う場合にも、異常な単語連鎖をもつ認識仮説は早期に棄却することが可能となるため、認識性能を低下させることなく認識処理速度を向上させることができるという効果が得られる。

【0057】なお、バイグラム確率値を0とおいた場合のように、計算機上ではその対数値が求められない場合には、バイグラム確率として0に近いきわめて小さな値を与えてもよく、上記と同様の効果を奏することはいう

12

*【0052】次に、言語尤度計算手段13は図2のステップST16において、各認識候補にて単語が遷移する際に、言語モデル階層決定手段12の決定した最適な階層を用いて、階層言語モデル記憶手段10に記憶されている階層化された統計的言語モデルから言語尤度を計算し、その対数値である対数言語尤度と、音響尤度計算手段5の計算した音響尤度の対数値である対数音響尤度から対数総合尤度を計算する。すなわち、単語連鎖x-wについて、選択された最適な階層を用いて、以下の式（8）のように言語尤度を計算する。なお、この式（8）において、 α は適当なスムージングのための係数である。

$$* \quad \text{【0053】} \\ \dots \dots \dots (8)$$

までもない。

【0058】実施の形態2. 次に、この発明の別の実施の形態について説明する。上記実施の形態1による音声認識装置における階層・クラス定義のような分類は、人間にとって直感的であるが外部知識が必要となる。この発明の実施の形態2はその外部知識を不要とし、各階層のクラスを先見的な知識なしに自動的に作成するものである。

【0059】図7はそのようなこの発明の実施の形態2による音声認識装置の構成を示すブロック図である。図において、1は音声取得手段、2は音響分析手段、3は音韻モデル記憶手段としてのHMM音韻モデル記憶手段、4は単語辞書記憶手段、5は音響尤度計算手段、8は認識結果出力手段、9はコーパス、12は言語モデル階層決定手段、13は言語尤度計算手段であり、これらは図1に同一符号を付して示した実施の形態1のそれらに相当する部分である。

【0060】また、14はあらかじめ定義された、統計的言語モデルの階層数を記憶する階層数定義記憶手段であり、15はこの階層数定義記憶手段14が記憶している階層数に基づいて、統計的言語モデルの階層構造を持つクラスを、コーパス9以外の外部知識を用いることなく、各階層においてそれぞれクラスタリングするクラス作成手段である。10は図1に同一符号を付して示した実施の形態1のそれに相当する、統計的言語モデルを記憶している言語モデル記憶手段としての階層言語モデル記憶手段であるが、上記クラス作成手段15にてクラスタリングされた階層のクラスと、コーパス9とを用いて作成した統計的言語モデルを記憶している点で、実施の形態1におけるそれとは異なっている。

【0061】次に動作について説明する。ここで、図8

はこの実施の形態2における、ある階層*i*でのクラス化処理の概略動作を示すフローチャートである。また、図9はある階層*i*での要素連鎖の、コーパス9での出現頻度を行列形式で示した説明図であり、最下層では要素は単語となり、それ以上の階層では要素はクラスとなる。さらに、図10はこの図9に示した行列を小行列に分解した例を示した説明図であり、図11は図9に示した行列について、その行と列を入れ替える操作を行った後の状態例を示した説明図である。

【0062】次に、この実施の形態2による音声認識装置のある階層でのクラス化の処理について説明する。以降は誤解を避けるために最下層におけるクラス化について説明する。なお、図9において、網掛けを施した部分はコーパス9の出現頻度が0より大きい数値であることを示し、白い部分はコーパス9の出現頻度が0であることを示している。

【0063】階層数定義記憶手段14はクラス言語モデルの階層数（例えば5）のみをあらかじめ定義して記憶す

$E = \text{小行列内要素数} / 0 \text{ 行列の小行列の個数}$

【0066】図10では、行列全体を4×4個の小行列に分割した場合に、その分割の仕方の一例を示している。ここで、連鎖の前後に同じクラス分けを行う場合には、行の分割の仕方と列の分割の仕方は同じでなければならず、したがって行列の対角線上にある小行列は図のように正方向行列となる。図10の場合、0行列である小行列は、最上段の右から2つ目にある4×3の小行列ただ1つであり、 $E = 12$ となる。なお、この行列のすべての分割の評価関数Eの最大値を E_{max} とする。

【0067】次に、ステップST42において任意の*i*と*j*（ $j \neq i$ ）とを選択し、ステップST43においてその行 w_i と行 w_j とを入れ替える操作を行い、ステップST44において列 v_i と列 v_j とを入れ替える操作を行う。次にステップST45において、入れ替えた後の行列に対して評価関数Eの最大値 E_{max} を求め、ステップST46において、その求められた E_{max} と以前の E_{max} とを比較する。その結果、求められた E_{max} が以前の E_{max} より大きければ、この方が精度のよいクラスタリングを与える行列表現であると考えて、ステップST47でもとの行列を入れ替え操作をした行列に更新し、 E_{max} も新たに求められたものに更新する。もし求められた E_{max} が以前の E_{max} より小さければ、以前の方が精度のよいクラスタリングを与える行列表現であると考えて、ステップST48では E_{max} を更新せずに行列をもとに戻す。

【0068】次に、ステップST49において収束条件に達したか否かの判定を行い、収束条件を満たすまで上記の操作を繰り返す。なお、収束条件としては、例えば次のようなものを用いることができる。

(1) 入れ替え操作の処理回数が所定数を越えた場合。

*しておく。クラス作成手段15はまず、ステップST40において、ある階層での要素（単語）連鎖の出現頻度を図9のような行列形式で表現する。次にその行列上で、コーパス9の出現頻度が0でない部分（網掛けを施した部分）と0である部分（白い部分）とを局在化させる処理を行う。0である部分と0でない部分が十分に局在化すれば、それはコーパス9に表れる連鎖と表れない連鎖をうまく分離できており、クラス化が適当であると考えることができる。

【0064】クラス作成手段15はステップST41において、その具体的な処理のために評価関数Eを導入して当該評価関数Eの計算を行う。この評価関数Eとしては、例えば、以下に示す式(9)を定義する。この式(9)は行列全体をN×N個の小行列に分割したときのものであり、Nはその階層に定義するクラスの数に相当する。なお、この式(9)中の小行列内要素数とは小行列が0行列のときのその面積である。

【0065】
$$E_{max} \dots (9)$$

(2) E_{max} が所定値を超えた場合。

【0069】以上の処理を各階層について行い、統計的言語モデルを作成した後、認識処理を行う。なお、認識処理は実施の形態1の場合と同一なので、ここではその説明を省略する。

【0070】ここで、評価関数の定義、収束条件の定義は、この実施の形態2に示したもののみに限定されるものでないことはいうまでもない。

【0071】また、この実施の形態2では単語連鎖（バイグラム）を2次元行列で説明したが、トライグラムでは3次元行列、一般のNグラムではN次元行列を用いれば、この実施の形態2をそのまま適用できることはいうまでもない。さらに、この実施の形態2におけるクラスタリング手法を先験的な知識を用いる実施の形態1におけるクラスタリングと組み合わせ、ある階層についてはこの実施の形態2のクラスタリング手法を、別の階層については先験的な知識を用いる実施の形態1のクラスタリング手法を用いることも可能であることはいうまでもない。

【0072】以上のように、この実施の形態2によれば、各階層をクラスタリングする際に先験的な知識がなくても、スパースなコーパス9についてクラスタリングすることが可能となるという効果が得られる。

【0073】

【発明の効果】以上のように、この発明によれば、音声データを所定時刻ごとに音響分析した音響特徴ベクトル、音韻モデルおよび単語辞書から計算した認識仮説の音響尤度と、言語モデル階層決定手段の決定した言語モデルの階層から統計的言語モデルを参照して計算した認識仮説の言語尤度に基づいて最終的な認識候補を出力するようにし、その統計的言語モデルに階層構造をもた

せ、音声認識の処理状況に応じて適切な階層の言語モデルを使用するように構成したので、コーバスの量が十分ではない場合でも、言語的につながり得ない異常な単語連鎖についての言語尤度を小さくすることが可能となるため、認識性能および認識処理速度の高い音声認識装置が得られるという効果がある。

【0074】この発明によれば、各先行単語によって最適な階層の言語モデルを使用するように構成したので、異常な単語連鎖に対する言語尤度が小さくなって、認識性能および認識処理速度の向上を図ることができるという効果がある。

【0075】この発明によれば、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーバスの出現頻度が0の場合、その単語バイグラムの確率値を0または0に近い小さな正値とするように構成したので、異常な単語連鎖に対する言語尤度が小さくなって、認識性能および認識処理速度の向上が図れるという効果がある。

【0076】この発明によれば、先行単語と、後続単語が属している適切な階層のクラスにおける単語-クラス連鎖のコーバスの出現頻度が0より大きく、かつ、コーバス内の単語バイグラムの出現頻度が0の場合には、その単語バイグラムの確率値を0より大きい値に設定するように構成したので、言語尤度や総合尤度の対数値を確実に計算できる効果がある。

【0077】この発明によれば、統計的言語モデルにおける先行単語に使用する最適な階層として、所属するクラスの単語バイグラムの出現頻度が0でない後続単語の数の変化率が最大になる直前の階層を選択するように構成したので、容易に最適な階層を決定することが可能になるという効果がある。

【0078】この発明によれば、コーバス以外の外部知識を用いて階層構造をもつクラスの作成を行うように構成したので、各階層のクラスについて人間にとって直感的な分類を行うことが可能になるという効果がある。

【0079】この発明によれば、外部知識を用いずに、各階層においてそれぞれクラスタリングをすることにより階層構造をもつクラスを作成するように構成したので、各階層のクラスを先験的な知識なしに自動的に作成することが可能になるという効果がある。

【0080】この発明によれば、ある一部の階層についてはコーバス以外の外部知識を用い、他の階層については外部知識を用いずにクラスタリングをすることにより、階層構造をもつクラスを作成するように構成したので、必要な階層では人間にとって直感的な分類が行え、他の階層では自動作成が可能になるという効果がある。

【0081】この発明によれば、クラスタリングすべき単語を行列表現して、行と列の入れ替えによってコーバスの出現頻度が0でない部分を局在させ、その部分の外

接するような部分行列をクラスとするように構成したので、コーバスに表れる連鎖と表れない連鎖とをうまく分離することができ、スパースなコーバスについてのクラスタリングを容易に行うことが可能になるという効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による音声認識装置の構成を示すブロック図である。

【図2】 実施の形態1による音声認識装置の動作を示すフローチャートである。

【図3】 階層構造をもつ単語辞書の一例を示す説明図である。

【図4】 階層構造をもつ言語モデルの、先行単語と後続単語の各階層のコーバス内での出現頻度を示す説明図である。

【図5】 実施の形態1における言語モデル階層決定手段の動作を示すフローチャートである。

【図6】 実施の形態1における言語モデル階層決定手段が計算した配列をグラフで示した説明図である。

【図7】 この発明の実施の形態2による音声認識装置の構成を示すブロック図である。

【図8】 実施の形態2におけるクラス作成手段の動作を示すフローチャートである。

【図9】 ある階層における要素連鎖のコーバスでの出現頻度を行列形式で示した説明図である。

【図10】 図9に示した行列を4×4の小行列に分解した例を示す説明図である。

【図11】 図9で示した行列について、その行と列を入れ替える操作を行った後の状態例を示した説明図である。

【図12】 この発明と従来の音声認識装置で用いられる単語辞書の具体例を示す説明図である。

【図13】 この発明と従来の音声認識装置における単語のHMMの構造の一例を示す説明図である。

【図14】 この発明と従来の音声認識装置にて、時刻が進むにつれて認識仮説が展開される状況を示す説明図である。

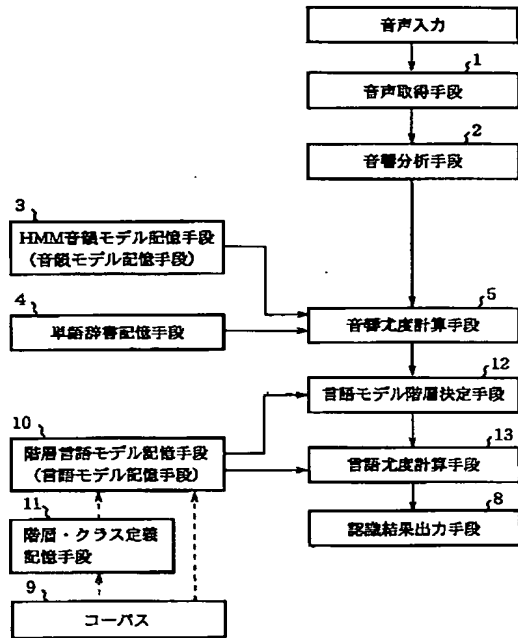
【図15】 従来の音声認識装置の構成を示すブロック図である。

【図16】 従来の音声認識装置の動作を示すフローチャートである。

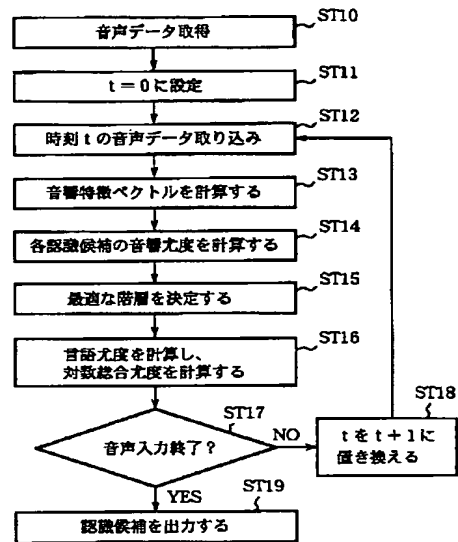
【符号の説明】

1 音声取得手段、2 音響分析手段、3 HMM音韻モデル記憶手段（音韻モデル記憶手段）、4 単語辞書記憶手段、5 音響尤度計算手段、8 認識結果出力手段、9 コーバス、10 階層言語モデル記憶手段（言語モデル記憶手段）、12 言語モデル階層決定手段、13 言語尤度計算手段。

【図1】



【図2】

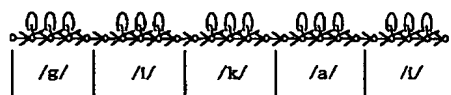


【図4】

【図3】

名詞						クラスレベル
医療用語						C4
薬品						C3
疾病						C2
...						C1
塩布薬	飲み薬	...	腹痛	C0
バンデリン	マイソジン	...	胃潰瘍	
アズノール	マイコシリン	
...	

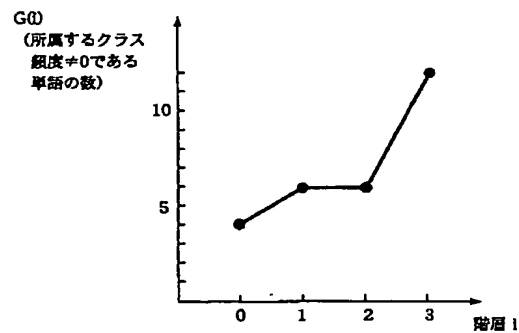
【図13】



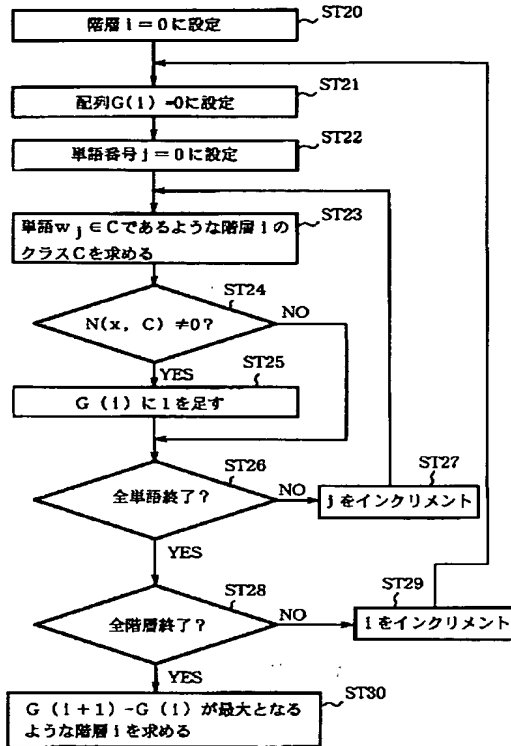
先行単語 x	54												クラスレベル
	0						54						
	0	0	0	0	0	0	12	5	32	0	5	0	
w ₁	w ₂	w ₃	w ₄	w ₅	w ₆	w ₇	w ₈	w ₉	w ₁₀	w ₁₁	w ₁₂		

後続単語 w

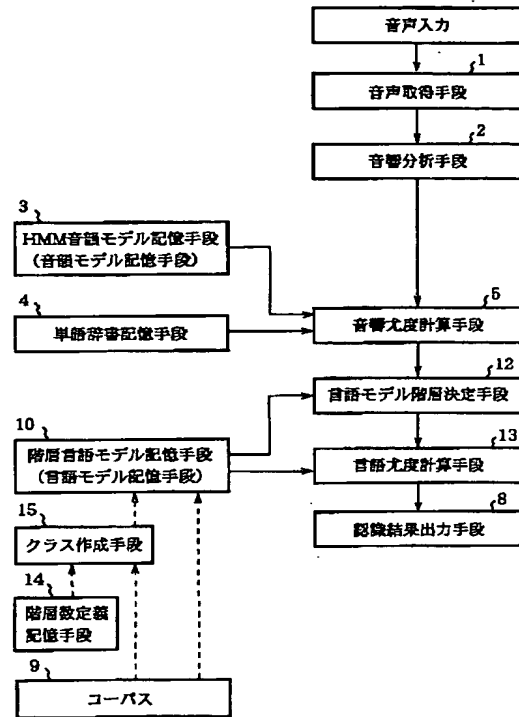
【図6】



【図5】

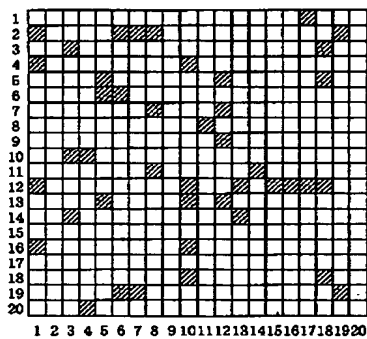


【図7】



【図9】

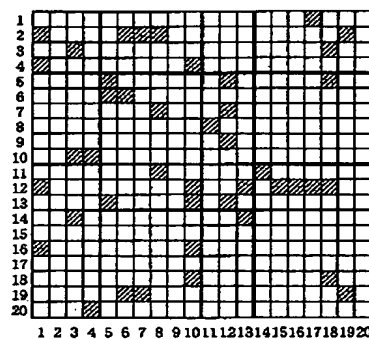
連続の前の単語番号



連続の後の単語番号

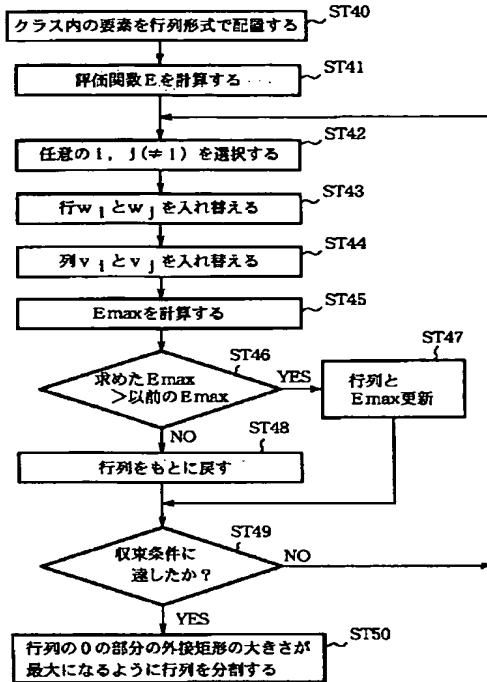
【図10】

連続の前の単語番号



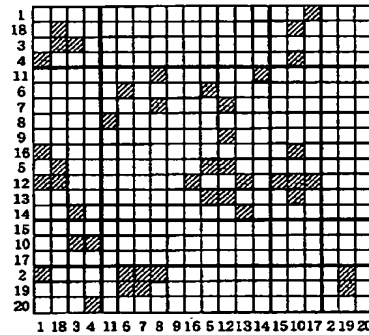
連続の後の単語番号

【図8】



【図11】

連鎖の前の単語番号

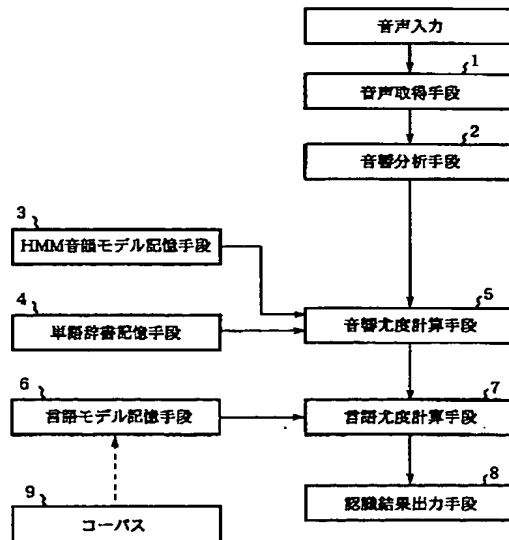


連鎖の後の単語番号

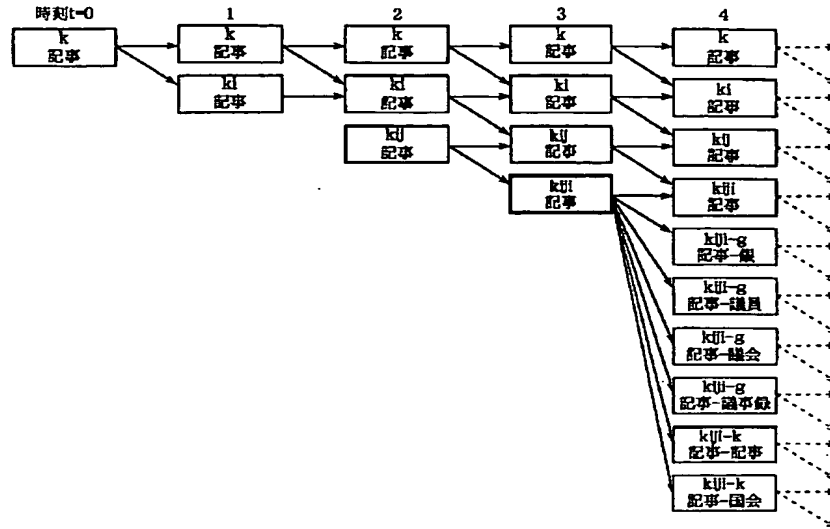
【図12】

漢字表記	ひらがな表記	音楽表記
...
銀	ぎん	giN
議員	ぎいん	giIN
議会	ぎかい	giKai
随事録	ぎじろく	giZiroku
記事	きじ	kiJI
国会	こっかい	koQkai
...

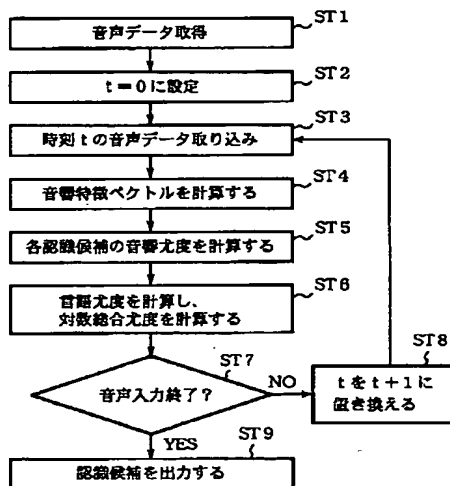
【図15】



【図14】



【図16】



This Page Blank (uspto)